UNCLASSIFIED

AD NUMBER

AD837009

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution authorized to U.S. Gov't. agencies and their contractors; Critical Technology; JUL 1968. Other requests shall be referred to Department of the Army, Fort Detrick, Attn: Technical Release Branch/TID, Frederick, MD 21701.

AUTHORITY

BDRL D/A ltr, 22 Oct 1971

THIS PAGE IS UNCLASSIFIED

AD837009

TRANSLATION NO. 993

DATE: 1 July 1968

DDC AVAILABILITY NOTICE

Reproduction of this publication in whole or in part is prohibited. However, DDC is authorized to reproduce the publication for United States Government purposes.

23

UNITED STATES ARMY
CHEMICAL CORPS BIOLOGICAL LABORATORIES
Fort Detrick, Maryland

Misc Tr 493

Tests and confidence intervals for exponential distributions, and their application to some discrete distributions

by J. Pfanzagl.

Translated from Metrika 3: 1: 1-25 (1960) by the Technical Library, Technical Information Division.

---

1. A general remark concerning randomized tests and confidence ranges.

Let there be a class of distributions P_θ which depend on a parameter θ. Let X be the appropriate random variable, and x its realizations. Further, let there be a definition of a family of null hypotheses $H_0(\delta): \theta \epsilon \omega_0(\delta)$. Each null hypothesis $H_0(\delta)$ is assumed to have a specific alternate hypothesis $H_1(\delta): \theta \epsilon \omega_1(\delta)$.

The problem of constructing a test for a specific null hypothesis $H_0(\delta_0)$, which in a certain sense is optimal for the appropriate alternative hypothesis $H_1(\delta_0)$, with due consideration for a prescribed safety factor α, frequently leads to a "criterion" $K(x, \delta_0)$, on the basis of which decisions about acceptance or rejection of the null hypothesis are made according to the following method:

$H_0(\delta_0)$ is accepted, if $K(x, \delta_0) < c$
$H_0(\delta_0)$ is rejected with probability p, if $K(x, \delta_0) = c$
$H_0(\delta_0)$ is always rejected, if $K(x, \delta_0) > c$.

The constants c and p must be chosen so that the prescribed safety factor α is maintained.

Let us assume that the distribution of $K(X, \delta_0)$ upon validity of $H_0(\delta_0)$ depends only on δ_0 (but not on $\theta \epsilon \omega_0(\delta_0)$). We now define the following quantity:

$$S_\delta(x, r) = P_\delta(K(X, \delta) < K(x, \delta)) + r\, P_\delta(K(X, \delta) = K(x, \delta)).$$

The function $S_\delta(x,r)$ permits construction of tests and confidence ranges in a very simple manner:

In order to bring clarity into the following considerations, we shall dispense with parameter δ , which at this moment is irrelevant, and write:

$$S(x,r) = P(K(X) < K(x)) + r\,P(K(X) = K(x)).$$

We now define a quantity c_α for each number $\alpha \in (0,1)$ by:

$$c_\alpha = \inf\{c : P(K(X) \leqq c) \geqq \alpha\}$$

Let R be a random variable equally distributed in $[0,1]$ which is independent of X . In this case:

1. $P^R(S(x,R) < \alpha) = 1$, if $K(x) < c_\alpha$

   Proof: For all $r \in [0,1]$, $S(x,r) \leqq P(K(X) \leqq K(x)) < \alpha$ is valid, if $K(x) < c_\alpha$

2. $P^R(S(x,R) < \alpha) = \dfrac{\alpha - P(K(X) < c_\alpha)}{P(K(X) = c_\alpha)}$, if $K(x) = c_\alpha$ and $P(K(X) = c_\alpha) > 0$.

   Proof: $P^R(P(K(X) < c_\alpha) + R\,P(K(X) = c_\alpha) < \alpha) =$

   $$= P^R\left(R < \frac{\alpha - P(K(X) < c_\alpha)}{P(K(X) = c_\alpha)}\right) = \frac{\alpha - P(K(X) < c_\alpha)}{P(K(X) = c_\alpha)},$$

   since $P(K(X) < c_\alpha) \leqq \alpha \leqq P(K(X) \leqq c_\alpha)$ is valid because of the left-sided continuity of $P(K(X) < c)$ and the right-sided continuity of $P(K(X) \leqq c)$.

3. $P^R(S(x,R) < \alpha) = 0$, if $K(x) > c_\alpha$

   Proof: For all $r \in [0,1]$, $S(x,r) \geqq P(K(X) < K(x)) \geqq \alpha$ is valid, if $K(x) > c_\alpha$.

   We now demonstrate that $S(X,R)$ is equally distributed in $[0,1]$: $P^{X,R}(S(X,R) < \alpha) = \alpha$.

   Proof: $P^{X,R}(S(X,R) < \alpha) = E^X P^R(S(X,R) < \alpha) =$

   $$= P(K(X) < c_\alpha) + \frac{\alpha - P(K(X) < c_\alpha)}{P(K(X) = c_\alpha)} \cdot P(K(X) = c_\alpha) = \alpha .$$

It follows from the preceding that the test method described above may also be formulated as follows: A realization (x, r) is determined and the null hypothesis is rejected, if $S(x,r) > \alpha$, otherwise it is accepted. This method guarantees that the null hypothesis is accepted (provided the prescribed safety factor α is observed), if $K(x) < c_\alpha$, that it is rejected with probability $p_\alpha = 1 - \frac{\alpha - P(K(x) < c_\alpha)}{P(K(x) = c_\alpha)}$ if $K(x) = c_\alpha$, and is always rejected, if $K(x) > c_\alpha$.

The function $S_\delta(x, r)$ also facilitates description of the confidence range which belongs to the test family discussed above: The quantity of that δ, for which $S_\delta(x,r) < \alpha$, contains the correct parameter value δ with probability α and thus yields a confidence range with safety factor α.

If $S_\delta(x,r)$ is a monotonic function of δ, the confidence range is even more easily defined: One simply determines the value of

$$\delta_\alpha = \inf \{ \delta : S_\delta(x,r) < \alpha \}$$

The confidence range therefore consists of all values $\delta > \delta_\alpha$ (if we assume that S is a monotonic decreasing function of δ).

In this case we may further use function S to obtain a median unbiased estimate function for δ. This function is given by $\delta_{\frac{1}{2}}$ (9, page 83).

These introductory remarks are designed to simplify notation in subsequent chapters. We shall generally limit ourselves to the "criterion" $K(x, \delta)$ applicable to each problem. Tests, confidence ranges and median unbiased estimate functions are derived from this criterion mechanically with the aid of function $S_\delta(x, r)$; for this reason the procedure itself is never included.

Since there are certain practical reservations against randomized methods, the author felt justified in working with the (non-randomized) function $S_\delta^*(x) = S_\delta\left(x, \frac{1}{2}\right)$. In repeated application to various test problems this method will surely produce the safety factor α on the average. At any rate, employment of function $S_\delta^*(x)$ would be preferable -- even from the practical point of view -- to the customary method which corresponds to the use of function $S_\delta(x, 0)$ and therefore yields safety factors of excessive size systematically (and, consequently, confidence intervals that are too wide systematically).

Next, a case of practical importance will be stressed, in which $S_\delta(x, r)$ is a monotonic decreasing function of δ:

Lemma 1: If the random variable $Y = k(x, \delta)$ contains a monotonic increasing density quotient, and if $K(x, \delta)$ is a monotonic decreasing (or constant) function of δ, then $S_\delta(x, r)$ is a monotonic decreasing function of δ.

Proof: It is known (9, page 74) that $\Psi(y)$ is a monotonic decreasing function of δ for every monotonic decreasing function associated with the monotonic increasing density quotient $E_\delta \Psi(Y)$. Therefore,

$$\Psi(y, \delta) = \begin{cases} 1 \text{ is valid for } y < K(x, \delta) \\ r \text{ is valid for } y = K(x, \delta) \\ 0 \text{ is valid for } y > K(x, \delta) \end{cases}$$

$E_\delta \Psi(Y, \delta) \geq E_{\delta'} \Psi(Y, \delta)$ is valid for $\delta < \delta'$.

However, in view of $K(x, \delta) \geq K(x, \delta')$, $\Psi(y, \delta) \geq \Psi(y, \delta')$ is also valid for $\delta < \delta'$, therefore:

$$E_{\delta'} \Psi(Y, \delta) \geq E_{\delta'} \Psi(Y, \delta')$$

and: $E_\delta \Psi(Y, \delta) \geq E_{\delta'} \Psi(Y, \delta')$ is valid for $\delta < \delta'$.

However, since $E_\delta \Psi(Y, \delta) = S_\delta(x, r)$ is valid for function $\Psi(y, \delta)$ defined above, the lemma is proved.

2. Exponential distributions.

Let us consider a class of distribution functions whose density with respect to an σ-finite measure μ defined in Euclidean space is represented by:

(1) $\quad p(x, \theta) = C(\theta) \exp[\theta x]$

where $\int p(x, \theta)\, d\mu(x) = 1$ is valid for $\theta \in \Omega$.

The study of such exponential distributions is of particular interest because an arbitrary class of distributions having a function $t(x)$, which is exhaustive with respect to its parameter, may always be reduced to the form:

$$p(x, \theta) = C(\theta) \exp[\theta t(x)]\, h(x)$$

by appropriate transformation of the parameter, provided a certain regularity exists and the range in which the densities are positive is independent of θ (8). Since $t(x)$ is exhaustive with respect to θ, we may restrict ourselves to the distribution of t. The latter has density $C(\theta)\exp[\theta t]$ with respect to an appropriate measure $\mu(t)$, which leads the general problem back to form (1).

Thus the results obtained from a study of distributions (1) are applicable to a large class of distributions with exhaustive functions. We shall subsequently discuss their application to certain discrete distributions.

3. Criteria for parameter θ.

Theorem 1: The criterion

$$K(x, \theta_0) = x$$

defines a most powerful test for the null hypothesis $\theta = \theta_0$ with respect to all alternative hypotheses $\theta > \theta_0$.

Proof: For $\theta > \theta_0$, the quotient of densities $\frac{p(x,\theta)}{p(x,\theta_0)}$ is a monotonic increasing function of x. For this reason, the above criterion leads, according to Theorem 3.2 in Lehmann (9, page 68) to a uniformly most powerful test in the range $\theta > \theta_0$.

We shall now consider the null hypothesis $H_0 : \theta = \theta_0$ and the alternative hypotheses $H_1 : \theta = \theta_0 - \Delta$ and $H_2 : \theta = \theta_0 + \Delta$. There is no test that would be most powerful with respect to both alternative hypotheses. We must, therefore, resort to another concept of optimality. Assuming that each of the two alternative hypotheses H_1 and H_2 has the same probability if H_0 is not valid, and that upon rejection of H_0 a classification (i.e., a decision for H_1 or H_2) must take place, we may use the optimal test and classification method developed in (11, page 15). The method is based on the following criterion:

$$K(x, \theta_0) = \mathrm{Max}\left[\frac{p(x,\theta_0-\Delta)}{p(x,\theta_0)}, \frac{p(x,\theta_0+\Delta)}{p(x,\theta_0)}\right] =$$

(2)

$$= \mathrm{Max}\left[\frac{C(\theta_0-\Delta)}{C(\theta_0)}\exp[-\Delta x], \frac{C(\theta_0+\Delta)}{C(\theta_0)}\exp[\Delta x]\right]$$

On the basis of this criterion we must define a function $S(x,r)$ and reject H_0 if $S_{\theta_0}(x,r) > \alpha$. The decision is then made in favor of H_1 or H_2, whichever is associated with a larger quotient of density functions. (If both quotients are identical, the decision is made arbitrarily).

A test and classification method constructed on the criterion defined by (2) guarantees that the probability of a correct classification is maximal in the event H_0 is invalid, provided the prescribed safety factor is observed (11, page 14,15).

Criterion (2) also depends on quantity Δ. If our interest is focussed principally on alternative hypotheses near the null hypothesis, we may employ a locally optimal test and classification method. We obtain the latter by using criterion

$$K'(x,\theta) = \left.\frac{\partial K(x,\theta)}{\partial \Delta}\right|_{\Delta=0}$$

Since

$$(3) \qquad \frac{c'(\theta)}{c(\theta)} = -E_{\theta}(X),$$

$$\left.\frac{\partial K(x,\theta)}{\partial \Delta}\right|_{\Delta=0} = \operatorname{Max}\left[-x+E_{\theta}(X), x-E_{\theta}(X)\right]$$

is valid. This proves:

Theorem 2: The criterion

$$K'(x,\theta_0) = |x - E_{\theta_0}(X)|$$

defines a locally optimal test and classification method for null hypothesis $\theta = \theta_0$ at point $\Delta = 0$ with respect to alternative hypotheses $H_1: \theta = \theta_0 - \Delta$ and $H_2: \theta = \theta_0 + \Delta$.

Generally speaking, the two-sided problem described here does not require delineation of the critical region in such a manner that both parts of the critical region have the same probability, but demands that they be symmetrical about the expected value. We may, of course, find a (bilaterally bounded) confidence interval for θ on the basis of this criterion.

Another possibility of obtaining a test which is, in a certain sense, optimal for hypothesis H_0 with respect to alternatives H_1 and H_2, is offered by limitation to universally unbiased tests. The class of unbiased tests contains a test which is most powerful with respect to both alternatives. This test still depends on Δ. Here, again, a locally most powerful test may be constructed when alternatives close to the null hypothesis are involved. The test obtained in this manner is found in Lehmann (9, page 126, 127). It is not identical with the criterion formulated in Theorem 2.

4. The 2-sample problem: Criteria for the difference of two parameters.

Let there be two distributions with densities

$$p_i\,(x, \theta_i) = C_i\,(\theta_i)\exp\,[\theta_i x], \qquad i = 1,2$$

with reference to measures μ_1 or μ_2. We now replace θ_1 and θ_2 by two new parameters θ and δ :

$$(4) \qquad \begin{aligned} \theta_1 &= \theta + \delta \\ \theta_2 &= \theta \end{aligned}$$

and search for a test for the (combined) null hypothesis $\delta = \delta_0,\ \theta \in \Omega$, i.e., we look for a test for the difference of θ_1 and θ_2 without making assumptions of any kind about the magnitude of either parameter.

The combined density of (X_1, X_2) with respect to measure $\mu = \mu_1 \times \mu_2$ is:

$$p(x_1, x_2, \theta, \delta) = C_1(\theta+\delta)\,C_2(\theta)\exp\left[\theta(x_1 + x_2) + \delta x_1\right].$$

We now consider the conditional distribution of $X_1 \mid X_1 + X_2 = x$.

According to Lemma 2.8 in Lehmann (9, page 52), there is a measure $\nu_x(x_1)$ which is such that the conditional distribution of $X_1 \mid X_1 + X_2 = x$ has the following density with respect to this measure:

$$(5) \qquad C_x(\delta)\exp\left[\delta x_1\right].$$

This density is independent of θ (i.e., function $x = x_1 + x_2$ is exhaustive with regard to parameter θ). Furthermore, the same lemma provides that the distribution of X (for stabilized δ) is an exponential distribution with the parameter θ, and therefore, according to Theorem 4.1 in Lehmann (9, page 132), complete for every value of δ. Consequently, based on a familiar theorem of Lehmann and Scheffe (cf. Theorem 4.2 in 9, page 134), every similar test for the above-mentioned null hypothesis has a Neyman structure, i.e., it is a conditional test for $X_1 + X_2 = x$. Thus the problem of finding a similar test for the above-named combined null hypothesis is reduced to the problem of finding a test for null hypothesis $\delta = \delta_0$ based on density (5). This problem was solved in the preceding chapter. We proved:

Theorem 3: The criterion

$$K(x_1, x_2, \delta_0) = x_1$$

defines a test for the combined null hypothesis $\delta = \delta_0, \theta \in \Omega$ which is most powerful similar with reference to all alternative hypotheses $\delta > \delta_0, \theta \in \Omega$, if we base our construction of function S on the conditional distribution of $X_1 \mid X_1 + X_2 = x$ (Formula 5).

Similarly, we may derive the corresponding two-sided problem from Theorem 2:

Again we start with representation (4) of parameters θ_1 and θ_2 by θ and δ, and obtain:

Theorem 4: The criterion

$$K'(x_1, x_2, \delta_0) = \left| x_1 - E_{\delta_0}(X_1 | x) \right|$$

defines a test and classification method for the null hypothesis $\delta = \delta_0, \theta \in \Omega$ which is local optimal similar at point $\Delta = 0$ with respect to alternative hypotheses $H_1 : \delta = \delta_0 - \Delta, \theta \in \Omega$ and $H_2 : \delta = \delta_0 + \Delta, \theta \in \Omega$, if we base our construction of function S on the conditional distribution of $X_1 \mid X_1 + X_2 = x$ (Formula 5).

5. Some versions of the k-sample problem.

Let there be k distributions with densities

$$p_i(x, \theta_i) = C_i(\theta_i) \exp[\theta_i x] \qquad i = 1, 2, \dots k$$

with respect to measures $\mu_1, \mu_2, \dots \mu_k$. We assume that the parameters $\theta_1, \theta_2, \dots \theta_k$ depend on two parameters θ and δ in the following form:

$$(6) \qquad \theta_i(\delta) = \vartheta + \Delta_i(\delta), \quad \text{with} \quad \Delta_i(0) = 0 \quad i = 1, 2, \dots k$$

Theorem 5: The criterion

$$K(x_1, x_2, \dots x_k, 0) = \sum_1^k \Delta_i(\delta_1) x_i$$

defines a test for the combined null hypothesis $\delta = 0, \theta \in \Omega$ which is most powerful similar with respect to alternative $\delta = \delta_1, \theta \in \Omega$, if we base our construction of function S on the conditional distribution of $X_1, X_2, \dots$

$$X_k \mid \sum_1^k X_i = x$$

Corollary: In the special case $\Delta_i(\delta) = \Delta_i \delta$ we finally obtain the criterion

$$K(x_1, x_2, \dots x_k, 0) = \sum_1^k \Delta_i x_i$$

which is independent of δ_1, making the test most powerful with reference to all $\delta > 0$.

Particular interest from the practical point of view is due in two cases, where $\Delta_i = i$ for $i = 1, 2, \dots k$ and $\Delta_i = 1$ for $i = 1, \dots \ell$; $\Delta_i = 0$ for $i = \ell + 1, \dots k$.

The proof of Theorem 5 is obvious: The combined density of $(X_1, X_2, \dots X_k)$ with reference to measure $\mu = \mu_1 \times \mu_2 \times \dots \times \mu_k$ is:

$$p(x_1, x_2, \dots x_k; \theta, \delta) = \left(\prod_1^k c_i(\theta + \Delta_i(\delta))\right) \exp\left[\theta \sum_1^k x_i + \sum_1^k \Delta_i(\delta) x_i\right]$$

We now consider the conditional distribution of $X_1, X_2, \dots X_k \mid \sum_1^k X_i = x$.

According to Lemma 2.8 in Lehmann (9, page 52), there is a measure $\nu_x(x_1, x_2, \dots x_k)$ which is such that the conditional distribution of

$$X_1, X_2, \dots X_k \mid \sum_1^k X_i = x$$

with reference to this measure has the following density:

$$C_x(\delta) \exp\left[\sum_1^k \Delta_i(\delta) x_i\right] \tag{7}$$

This density is independent of θ. The same lemma provides that the distribution of x (for stabilized δ) is an exponential distribution with the parameter θ and therefore, according to Theorem 4.1 in Lehmann (9, page 132), complete. For this reason and according to the above-mentioned theorem of Lehmann and Scheffe (Theorem 4.2 in 9, page 134), every similar test for $\delta = 0, \theta \in \Omega$ has a Neyman structure, i.e., it is a conditional test for $X_1 + X_2 + \dots + X_k = x$. Consequently, if we construct the quotient of densities (7) for $\delta = \delta_1$ and $\delta = 0$, we see that it is a monotonic function of

$$\sum_1^k \Delta_i(\delta_1) x_i,$$

proving Theorem 5.

Let us consider another version of the k-sample problem: Let there be m hypotheses $H_1, \dots H_m$. Hypothesis H_j establishes the values $\Delta_{ij}(\delta)$ in a quite specific manner (where $\Delta_{ij}(0) = 0$ always applies to all i, j). We shall discuss the following cases in detail:

a) $\Delta_{ij}(\delta) = \begin{cases} \delta \text{ for } i = j \\ 0 \text{ for } i \neq j \end{cases} \qquad i, j = 1, 2, \dots k$

In this case hypothesis H_j indicates that the j^{th} distribution — and only this one — has a parameter value $(\theta+\delta)$ which deviates from the remaining distributions.

b) $\Delta_{ij}(\delta) = \begin{cases} \delta \text{ for } i \geq j \\ 0 \text{ for } i < j \end{cases} \qquad i, j = 1, 2, \dots k.$

In this case hypothesis H_j indicates that a jump had occurred at point j and that the distributions $j, j+1, \dots k$ have a parameter $(\theta+\delta)$ which deviates from the remaining distributions.

In treating the problem formulated above, we assume that each of the alternative hypotheses $H_1, \cdots H_m$ has the same probability, and that a decision must be made for one of the alternative hypotheses, if the null hypothesis $H_0: \delta = 0, \theta \in \Omega$ is invalid and must be rejected.

Theorem 6: The criterion

$$K(x_1, x_2, \dots x_k, 0) = \underset{j=1,\dots m}{\mathrm{Max}} \left[\sum_{i=1}^{k} \Delta'_{ij}(0)(x_i - E_0(X_i \mid x)) \right]$$

defines a test and classification method for the combined null hypothesis $\delta = 0, \theta \in \Omega$ which is local optimal similar at point $\delta = 0$ with respect to alternative hypotheses $H_1 \dots H_m$, if we base our construction of function S on the conditional distribution of $X_1, X_2, \dots X_k \mid \sum_{1}^{k} X_i = x$ (Formula 7).

Corollary a: In the special case a, we get the criterion:

$$\underset{j=1,\dots k}{\mathrm{Max}} \left[x_j - E_0(X_j \mid x) \right]$$

Incidentally, this criterion is uniformly optimal (for all $\delta > 0$) in case all measures $\mu_1, \mu_2, \dots \mu_k$ are identical, and, since

$$E_0(X_1 \mid x) = E_0(X_2 \mid x) = \dots = E(X_k \mid x),$$

is further simplified to:

$$\underset{j=1,\dots k}{\mathrm{Max}} \left[x_j \right]$$

Corollary b: In the special case b, we get as criterion:

$$\max_{j=1,\ldots k} \left[\sum_{i=j}^{k} (x_j - E_0 (X_i | x)) \right] \tag{8}$$

Since the proof of Theorem 5 was presented in detail, proof of Theorem 6 may be abbreviated. As above, it is shown that every similar test of the null hypothesis must be a conditional test. According to (11, page 15),

$$\max_{j=1,\ldots m} \left[\frac{p_j (x_1, \ldots x_k, \delta)}{p (x_1, \ldots x_k, 0)} \right]$$

is the optimal criterion of decision. Here $p_j (x_1, \ldots x_k, \delta)$ is the specific form assumed by density (7) under hypothesis H_j, therefore

$$p_j (x_1, \ldots x_k, \delta) = C_x^{(j)} (\delta) \exp \left[\sum_1^k \Delta_{ij} (\delta) x_i \right].$$

$p (x_1, \cdots x_k, 0)$ is the density in case of validity of the null hypothesis. In order to obtain a local optimal decision procedure at point $\delta = 0$, we merely employ the criterion

$$\max_{j=1,\ldots m} \left[\frac{p_j' (x_1, \ldots x_k, 0)}{p (x_1, \ldots x_k, 0)} \right]$$

where $p_j' (x_1, \ldots x_k, 0)$ is the derivative of $p_j (x_1, \ldots x_k, \delta)$ after δ at point $\delta = 0$.

Since
$$\frac{C_x^{(j)'} (0)}{C_x^{(j)} (0)} = - \sum_1^m \Delta_{ij}' (0) E_0 (X_i | x),$$

$$\frac{p_j' (x_1, \ldots x_k, 0)}{p (x_1, \ldots x_k, 0)} = \sum_1^m \Delta_{ij}' (0) (x_i - E_0 (X_i | x))$$

is valid, which proves our theorem.

Specialization for $k = 2$ leads us once again to the 2-sample problem:

Theorem 5 is converted to Theorem 3, since the quantity $\Delta_1 (\delta_1) x_1 + \Delta_2 (\delta_1) x_2$ is a monotonic function of x_1 for $x_1 + x_2 = x$.

Theorem 6 is converted to Theorem 4, if we make the obvious assumption that $\Delta_{11}(\delta)=\Delta_{22}(\delta)$ and $\Delta_{12}(\delta)=\Delta_{21}(\delta)$. In this case the criterion of Theorem 6 proves equivalent to $|x_1 - E_0(X_1|x)|$ because $x_1+x_2=x$.

6. The k-sample problem upon r repetitions.

Let there be $k \times r$ distributions with following densities with reference to certain measures μ_{ij} $(i=1,2,\ldots k; j=1,2,\ldots r)$:

$$p_{ij}(x,\theta_{ij}) = C_{ij}(\theta_{ij})\exp\left[\theta_{ij}x\right] \qquad \begin{matrix} i=1,2,\ldots k \\ j=1,2,\ldots r \end{matrix}$$

We assume that parameters θ_{ij} are dependent on certain parameters $(\theta_1, \theta_2, \ldots \theta_r, \delta)$ in the following manner:

$$\text{(9)} \quad \theta_{ij} = \theta_j + \Delta_i(\delta) \text{ with } \Delta_i(0)=0 \qquad \begin{matrix} i=1,2,\ldots k \\ j=1,2,\ldots r \end{matrix}$$

We are faced with the task of finding a similar test for the combined null hypothesis $H_0: \delta=0,\ \theta_j \in \Omega,\ j=1,2,\ldots r$.

In practice, such a problem occurs when a test is designed to establish whether k distributions are identical and the test is repeated r times, although it is impossible to arrange repetition such a way that θ_j has the same value for all repetitions. The formulation chosen above seems adequate when the situation promises, upon failure of the null hypothesis, that values $\Delta_1(\delta), \Delta_2(\delta), \ldots \Delta_k(\delta)$ are identical in all repetitions, e.g., that an increasing trend is present in all repetitions, or that one distribution deviates from the others, that this distribution is not identified from the start, although it is clear that the deviation would be valid for all repetitions and would always involve the same distribution.

As in the preceding chapters, we consider that each similar test must be a test with Neyman's structure. If we designate the realizations of distribution p_{ij} as x_{ij}, and $x_{i1}+x_{i2}+\ldots+x_{ir}=x_{i\cdot}$, while $x_{1j}+x_{2j}+\ldots+x_{kj}=x_{\cdot j}$, we may formulate the following theorems:

Theorem 7: The criterion

$$K(x_{11}, \ldots x_{kr}, 0) = \sum_{1}^{k} \Delta_i(\delta_1)\, x_{i\cdot}$$

defines a test for the combined null hypothesis $\delta=0,\ \theta_j \in \Omega,\ j=1,2,\ldots r$ which is most powerful similar with respect to the alternative $\delta=\delta_1,\ \theta_j \in \Omega,\ j=1,2,\ldots r$, if we base our construction of function S on the conditional distribution of $X_{11}, \ldots X_{kr} \mid X_{\cdot 1}=x_{\cdot 1}, \ldots X_{\cdot r}=x_{\cdot r}$

Corollary: In the special case $\Delta_i(\delta) = \Delta_i \delta$ we ultimately obtain the criterion

$$K(x_{11}, \ldots x_{kr}, 0) = \sum_{1}^{k} \Delta_i x_{i\cdot}$$

which is independent of δ_1, making the test most powerful with respect to all $\delta > 0$.

The proof is analogous to proof of Theorem 5.

Theorem 8: The criterion

$$K(x_{11}, \ldots x_{kr}, 0) = \underset{\ell = 1, \ldots m}{\mathrm{Max}} \left[\sum_{i=1}^{k} \Delta'_{i\ell}(0) \sum_{j=1}^{r} (x_{ij} - E_0(X_{ij} | x_{\cdot j})) \right]$$

defines a test and classification method for the combined null hypothesis $\delta = 0, \theta_j \in \Omega, j = 1, 2, \ldots r$ which is local optimal similar at point $\delta = 0$ with respect to alternative hypothesis $H_1, \ldots H_m$, if we base our construction of function S on the conditional distribution of $X_{11}, \ldots X_{kr} | X_{\cdot 1} = x_{\cdot 1}, \ldots X_{\cdot r} = x_{\cdot r}$.
(Here hypothesis H_ℓ determines the values $\Delta_i(\delta)$ with $\Delta_{1\ell}(\delta), \ldots \Delta_{k\ell}(\delta)$.
Cf. page 10 for this and special cases a and b).

Corollary a: In the special case a we get as criterion:

$$\underset{i = 1, \ldots k}{\mathrm{Max}} \left[\sum_{j=1}^{r} (x_{ij} - E_0(X_{ij} | x_{\cdot j})) \right]$$

Incidentally, this criterion is uniformly optimal (for all $\delta > 0$) in case measures $\mu_{1j}, \mu_{2j}, \ldots \mu_{kj}$ for constant j are identical, and is further simplified to

$$\underset{i = 1, \ldots k}{\mathrm{Max}} [x_{i\cdot}]$$

because $$E_0(X_{1j} | x_{\cdot j}) = E_0(X_{2j} | x_{\cdot j}) = \ldots = E(X_{kj} | x_{\cdot j})$$

Corollary b: In the special case b we get as criterion:

$$\underset{\ell = 1, \ldots k}{\mathrm{Max}} \left[\sum_{i=\ell}^{k} \sum_{j=1}^{r} (x_{ij} - E_0(X_{ij} | x_{\cdot j})) \right]$$

7. Some theorems pertaining to reproductive distributions.

Premise: Let there be a class of distributions which depends on two parameters θ and λ, and which is complete for every fixed λ. Further, let this class be reproductive for every fixed θ with respect to the parameter λ:

and let the sum variable $X = X_1 + X_2$ be exhaustive for all values with respect to parameter θ.

Let $X_i (i = 1, \dots, k)$ be independent random variables with distributions P_{θ, λ_i}. The sum variable X in that case has the distribution $P_{\theta, \lambda}$ with $\lambda = \sum_1^k \lambda_i$.

Lemma 2: On the premise defined above,

$$E(X_i | x) = \frac{\lambda_i}{\lambda} x.$$

Proof: Due to reproductivity,

$$E(X_i) = \frac{\lambda_i}{\lambda} E(X).$$

For this reason, $E^X(E(X_i|X)) = E(X_i) = \frac{\lambda_i}{\lambda} E(X) = E\left(\frac{\lambda_i}{\lambda} X\right)$,

that is, $E^X\left(E(X_i|X) - \frac{\lambda_i}{\lambda} X = 0\right.$.

According to the premise, X is exhaustive with respect to θ, and $E(X_i | x)$ is independent of θ. Furthermore, since X has distribution $P_{\theta, \lambda}$ and the latter is complete according to the premise,

$E(X_i | x) - \frac{\lambda_i}{\lambda} x = 0$ is valid for all x ($P_{\theta, \lambda}$ - nearly everywhere).

Lemma 3: On the premise defined above,

$$V(X_i | x) = \frac{\lambda_i}{\lambda}\left(1 - \frac{\lambda_i}{\lambda}\right) v(x, \lambda)$$

$$C(X_i X_j | x) = -\frac{\lambda_i}{\lambda} \cdot \frac{\lambda_j}{\lambda} v(x, \lambda).$$

$v(x, \lambda)$ is the unbiased estimate function for $V(x)$.

Proof: Due to reproductivity,

$$V(X_i) = \frac{\lambda_i}{\lambda} V(X).$$

Therefore, $E^X(V(X_i|X)) = E^X(E(X_i^2|X)) - E^X(E(X_i|X)^2) =$

$$= E(X_i^2) - \frac{\lambda_i^2}{\lambda^2}E(X^2) = V(X_i) + E(X_i)^2 - \frac{\lambda_i^2}{\lambda^2}E(X^2) =$$

$$= \frac{\lambda_i}{\lambda}V(X) + \frac{\lambda_i^2}{\lambda^2}E(X)^2 - \frac{\lambda_i^2}{\lambda^2}E(X^2) = \frac{\lambda_i}{\lambda}\left(1 - \frac{\lambda_i}{\lambda}\right)V(X),$$

and

$$E^X\left(\frac{V(X_i|X)}{\frac{\lambda_i}{\lambda}\left(1-\frac{\lambda_i}{\lambda}\right)}\right) = V(X).$$

Thus $\dfrac{V(X_i|x)}{\frac{\lambda_i}{\lambda}\left(1-\frac{\lambda_i}{\lambda}\right)}$ is an unbiased estimate function for $V(X)$

(independent of θ). However, since the distribution of X is complete for every value of λ , the unbiased estimate function for $V(X)$ ($P_{\theta,\lambda}$ - nearly everywhere) is clearly defined. If we designate the unbiased estimate function for $V(X)$ by $v(x,\lambda)$, then

$$V(X_i|x) = \frac{\lambda_i}{\lambda}\left(1-\frac{\lambda_i}{\lambda}\right)v(x,\lambda) \qquad (P_{\theta,\lambda} \text{ - nearly everywhere})$$

is valid and proves the ratio of variances. The ratio of covariances follows directly from

$$V(X_i+X_j|x) = V(X_i|x) + V(X_j|x) + 2\,C(X_iX_j|x)$$

Lemma 4:

$$E\left(\sum_1^k a_i X_i | x\right) = E[a]\,x$$

$$V\left(\sum_1^k a_i X_i | x\right) = V[a]\,v(x,\lambda)$$

Here

$$E[a] = \sum_1^k a_i\frac{\lambda_i}{\lambda},\quad V[a] = \sum_1^k (a_i - E[a])^2\frac{\lambda_i}{\lambda}.$$

The proof is derived directly from Lemmas 2 and 3.

8. Specialization of general theorems concerning exponential distributions for reproductive exponential distributions.

Although notation of density in form (1) and incorporation of any factors independent of θ in the measure μ facilitated general considerations, it is more feasible in application to let the measure remain fixed. This is particularly valid in case a second reproductive parameter λ exists in addition to θ. In this event we take the measure $\mu(x)$ as a basis, which is independent of λ and with respect to which $P_{\theta,\lambda}$ has density

$$C(\theta,\lambda)\exp[\theta,x]f(x,\lambda)$$

In this case the conditional distribution of $X_1, X_2, \dots X_k \mid \sum_1^k X_i = x$ has density

$$C_x(\delta;\lambda_1,\dots\lambda_k)\exp\left[\sum_1^k \Delta_i(\delta)x_i\right]\frac{\prod_1^k f(x_i,\lambda_i)}{f(x,\lambda)} \tag{10}$$

with respect to measure $\mu(x_1)\times\mu(x_2)\times\dots\times\mu(x_{k-1})$ under the conditions listed on page 13. (This formula replaces Formula 7, p. 9).

In case $\delta = 0$, this density becomes

$$\frac{\prod_1^k f(x_i,\lambda_i)}{f(x,\lambda)} \tag{11}$$

for all $i = 1,2,\dots k$ due to $\Delta_i(0) = 0$.

Lemma 2 permits further simplification of criteria defined in Theorems 4, 6 and 8 in case of reproductivity, and allows conversion to even more feasible forms under certain conditions. We shall dispense with their detailed treatment and merely list a few examples:

Theorem 4 (page 8) proposes $|x_1 - E_{\delta_0}(X_1|x)|$ as criterion. In case $\delta_0 = 0$ this criterion proves equivalent to

$$K''(x_1,x_2,0) = \left|\frac{x_1}{\lambda_1} - \frac{x_2}{\lambda_2}\right| \tag{12}$$

due to $E(X_1|x) = \frac{\lambda_1}{\lambda}x$, when $\lambda = \lambda_1 + \lambda_2$ and $x = x_1 + x_2$ are considered.

Furthermore, we may, under certain conditions, approximate the criterion's distribution by a normal distribution, if the null hypothesis is valid. This requires standardization. Upon application of Lemma 4, standardization of criterion (12) leads directly to form:

$$(13) \qquad \frac{\left| \frac{x_1}{\lambda_1} - \frac{x_2}{\lambda_2} \right|}{\sqrt{\frac{v(x, \lambda)}{\lambda_1 \lambda_2}}}$$

Upon standardization of the criterion $\sum_1^m \Delta_i x_i$ listed as corollary to Theorem 5 (page 8), Lemma 4 immediately supplies the standardized form of the criterion:

$$(14) \qquad \frac{\sum_1^k \Delta_i x_i - E[\Delta] x}{\sqrt{V[\Delta] v(x, \lambda)}} \quad \text{or} \quad \frac{\sum_1^k (\Delta_i - E[\Delta]) x_i}{\sqrt{V[\Delta] v(x, \lambda)}}$$

where $E[\Delta] = \sum_1^k \Delta_i \frac{\lambda_i}{\lambda}, \quad V[\Delta] = \sum_1^k (\Delta_i - E[\Delta])^2 \frac{\lambda_i}{\lambda}$.

Finally, we shall specialize Theorem 7 (page 12) for $k = 2$ in the case of reproductive distributions. As in Chapter 5, we consider that the criterion is simplified to:

$$K(x_{11}, \dots x_{kr}, 0) = x_1 .$$

in case $k = 2$. In case of reproductivity we get the criterion

$$(15) \qquad \frac{\sum_{j=1}^r \frac{\lambda_{1j} \lambda_{2j}}{\lambda_{\cdot j}} \left(\frac{x_{1j}}{\lambda_{1j}} - \frac{x_{2j}}{\lambda_{2j}} \right)}{\sqrt{\sum_{j=1}^r \frac{\lambda_{1j} \lambda_{2j}}{\lambda_{\cdot j}^2} v(x_{\cdot j}, \lambda_{\cdot j})}}$$

by standardization based on Lemma 4.

In practice, one frequently has a free choice of parameters $\lambda_1, \lambda_2, \dots \lambda_k$, provided certain limitations are observed (e.g., $\sum_1^k \lambda_i \leq \lambda_0$). In this paper we shall not discuss how parameters $\lambda_1, \lambda_2, \dots \lambda_k$ are to be chosen in order to afford precise information about θ.

9. Application to various discrete distributions.

There are various discrete distributions which are amenable to conversion to an exponential form through appropriate transformation of their parameter, and are, in addition, reproductive with respect to a second parameter. Specifically, these include:

Binomial distribution: $\binom{\lambda}{x} p^x (1-p)^{\lambda - x}$.

Transformation $\theta = \log \frac{p}{1-p}$ yields $(1 + \exp[\theta])^{-\lambda} \exp[\theta x] \binom{\lambda}{x}$.

Poisson distribution: $\frac{(\lambda a)^x}{x!} \exp[-\lambda a]$

This notation deviates from the customary style in that the mean value has been entered as the product of two parameters. The additional parameter λ proved feasible because many applications present a situation in which the mean values of distributions are related in certain known proportions (λ_i), while the common factor of proportionality (a) is unknown.

The transformation $\theta = \log a$ produces: $\exp[-\lambda \exp[\theta]] \exp[\theta x] \frac{\lambda^x}{x!}$.

Negative binomial distribution: $\binom{\lambda + x - 1}{x} p^x (1-p)^{\lambda}$.

The transformation $\theta = \log p$ yields: $(1 - \exp[\theta])^{\lambda} \exp[\theta x] \binom{\lambda + x - 1}{x}$

Pascal distribution: $\binom{x-1}{\lambda - 1} (1-p)^{x-\lambda} p^{\lambda}$.

The transformation $\theta = \log(1-p)$ yields: $(\exp[-\theta] - 1)^{\lambda} \exp[\theta x] \binom{x-1}{\lambda - 1}$.

It is known that all of these distributions are reproductive with respect to parameter λ, making them accessible to Lemmas 2, 3 and 4. As shown by an elementary computation, the function $v(x, \lambda)$ takes the following form in these four distributions:

Binomial distribution: $v(x, \lambda) = \frac{x(\lambda - x)}{\lambda - 1}$

Poisson distribution: $v(x, \lambda) = x$

Negative binomial distribution: $v(x, \lambda) = \frac{x(\lambda + x)}{\lambda + 1}$

Pascal distribution: $v(x, \lambda) = \frac{x(x - \lambda)}{\lambda + 1}$

Basically, we can now explicitly elaborate any one of Theorems 1-8 (when required, with due consideration of Lemmas 2 and 4) for each of the specific distributions listed above. Instead, we shall isolate certain special cases:

1. Propositions about the difference of two parameters.

In order to make unilateral propositions about the difference of two parameters, we must, according to Theorem 3 (page 8), depart from the criterion x_1 and consider the conditional distribution of $x_1 \mid x_1 + x_2 = x$. According to Formula 10 (special case $\Delta_1(\delta) = \delta$, $\Delta_2(\delta) = 0$), this conditional distribution possesses the density:

$$C_x(\delta, \lambda_1, \lambda_2) \exp[\delta x_1] \frac{f(x_1, \lambda_1) f(x - x_1, \lambda_2)}{f(x, \lambda)},$$

where $C_x(\delta, \lambda_1, \lambda_2) = \left(\int \exp[\delta x_1] \frac{f(x_1, \lambda_1) f(x - x_1, \lambda_2)}{f(x, \lambda)} d\mu(x_1) \right)^{-1}$

Specializing this general formula in the Poisson distribution, we get density:

$$\binom{x}{x_1} \left(\frac{\lambda_1 e^\delta}{\lambda_1 e^\delta + \lambda_2} \right)^{x_1} \left(1 - \frac{\lambda_1 e^\delta}{\lambda_1 e^\delta + \lambda_2} \right)^{x - x_1}$$

This is the density of a binomial distribution with parameter $p = \frac{\lambda_1 e^\delta}{\lambda_1 e^\delta + \lambda_2}$. In computing function S, we may therefore employ the tables of summable functions of the binomial distribution. In this manner we obtain a proposition with respect to δ. Even more clarity is obtained by converting this proposition about δ into a proposition about e^δ, since $e^\delta = \frac{a_1}{a_2}$ because $\theta = \log a$.

We thus obtain, as desired, a test, a confidence interval or a median unbiased estimate function for the quotient $\frac{a_1}{a_2}$. The test for hypothesis $\frac{a_1}{a_2} = 1$ (i.e., $a_1 = a_2$) was first defined by Przyborowski and Wilenski (12). Its optimality was proved by Tocher (13). Bross (2) employed these considerations in the construction of a confidence interval.

Specialization in the case of binomial distribution leads to hypergeometric distribution if $\delta = 0$. We thereby get R.A. Fisher's (5) test as optimal test for hypothesis $\delta = 0$ (i.e., $p_1 = p_2$). Its optimality was also demonstrated by Tocher (13). In the general case we get a test of an arbitrary value of δ (or a confidence interval or a median unbiased estimate of δ). If we convert this proposition about δ into a proposition about the original parameter p_1, p_2, we get a proposition about $\left(\log \frac{p_1}{1-p_1} - \log \frac{p_2}{1-p_2}\right)$ because $\theta = \log \frac{p_1}{1-p_1}$, or by antilogarithmic treatment, a proposition about $\left(\frac{p_1}{1-p_1} : \frac{p_2}{1-p_2}\right)$. This appears unwelcome at first glance, since one is accustomed to thinking primarily in values of p and a proposition about $(p_1 - p_2)$ or $\frac{p_1}{p_2}$ would be more desirable. Actually, neither $(p_1 - p_2)$ nor $\frac{p_1}{p_2}$ is a plausible measure for the difference between p_1 and p_2. An increase from 1% to 2% is not the same as an increase from 10% to 11% or from 50% to 51%. Nor is it the same as an increase from 10% to 20% or from 50% to 100%. On this basis, Koller (7, page 324) proposed — independently of the present considerations — the value $\left(\frac{p_1}{1-p_1} : \frac{p_2}{1-p_2}\right)$ as measure of the difference. Moreover, the transformation $\theta = \log \frac{p}{1-p}$ is widely accepted in connection with the evaluation of activity curves. It follows naturally from the assumption that the effectiveness of a drug (i.e., the portion p of those cases in which the effect was observed) depends on the dosage in form of a logistic function (cf., for example, Berkson 1). Finally, let me point out that Weichselberger concluded in (17) that the quantity $\left(\frac{p_1}{1-p_1} : \frac{p_2}{1-p_2}\right)$ is an adequate measure, in certain models, for the degree of dependence in a 2x2 contingency table.

2. A test against trend.

If Formula (14) (page 17) for $\Delta_i = i$ is specialized in the case of binomial distribution, the result is:

$$\frac{\sum_1^k (i - E[i]) x_i}{\sqrt{V[i] \frac{x(\lambda - x)}{\lambda - 1}}} \quad \text{with} \quad E[i] = \sum_1^k i \frac{\lambda_i}{\lambda}, \quad V[i] = \sum_1^k (i - E[i])^2 \frac{\lambda_i}{\lambda}.$$

Essentially the same criterion is obtained by specialization of the results of Yates (14, p. 178) in the case of a 2xk table. (The sole difference consists in the fact that Yates employs λ in the denominator in place of $\lambda - 1$). Yates' formula was adopted by Cochran (3, p. 435, Formula 14), where the reader finds practical examples.

Incidentally, this test shows a certain kinship with the test of v. Eeden and Hemelrijk (4), although it is by no means equivalent to the latter. Using our designations, the criterion employed by v. Eeden and Hemelrijk is:

$$\frac{\sum_1^k \left(i - \frac{k+1}{2}\right) \frac{x_i}{\lambda_i}}{\sqrt{\left(\sum_1^k \left(i - \frac{k+1}{2}\right)^2 \frac{1}{\lambda_i}\right) \frac{x(\lambda - x)}{\lambda(\lambda - 1)}}}$$

Aside from employment of the unweighted mean $\frac{k+1}{2}$ in place of $\sum_1^k i \frac{\lambda_i}{\lambda}$, an essential difference involves construction of the criterion on $\frac{x_i}{\lambda_i}$ instead of on x_i.

In the special case $\lambda_i = 1$ for $i = 1, 2, \ldots k$, x_i can only accept the values $x_i = 0$ and $x_i = 1$. Consequently the criterion $\sum_1^k i\, x_i$ which follows for $\Delta_i = i$ from the corollary to Theorem 5 (page 8) becomes identical with the sum of those rank numbers i for which $x_i = 1$. This criterion of rank sums was used by Haldane and Smith (6). It proved to be equivalent to Wilcoxon's (Mann-Whitney) test, if the quantity $\{i : x_i = 1\}$ is considered the rank numbers of one sample and the quantity $\{i : x_i = 0\}$ as the rank numbers of the second sample.

3. The combination of 2x2 contingency tables.

Specialization of Formula (15) (page 17) in the case of binomial distribution yields:

$$(16) \qquad \frac{\sum_{i}^{r} \frac{\lambda_{1j}\,\lambda_{2j}}{\lambda_{.j}} \left(\frac{x_{1j}}{\lambda_{1j}} - \frac{x_{2j}}{\lambda_{2j}} \right)}{\sqrt{\sum_{1}^{r} \frac{\lambda_{1j}\,\lambda_{2j}}{\lambda_{.j}^{2}} \cdot \frac{x_{.j}(\lambda_{.j} - x_{.j})}{\lambda_{.j} - 1}}}$$

Since the test for independence in a 2x2 contingency table can also be reduced to a comparison of two binomial distributions (cf., for example, Lehmann 9, page 143), we may apply this formula to the combination of r 2x2 contingency tables. It will be optimal if an existing dependence has the same direction in every contingency table and corresponds in quantity to a constant logit effect, i.e.,

$$\frac{p_{2j}}{1 - p_{2j}} = k \frac{p_{1j}}{1 - p_{1j}} \quad \text{for all } j = 1, 2, \dots r.$$

The custom of simply adding the χ^2 values of individual contingency tables is unfeasible in this case because the direction of dependence is not being considered, and the consistency of dependence in the various contingency tables does not come into play.

A somewhat refined criterion,

$$\frac{1}{\sqrt{r}} \sum_{j=1}^{r} \frac{\frac{x_{1j}}{\lambda_{1j}} - \frac{x_{2j}}{\lambda_{2j}}}{\sqrt{\frac{x_{.j}(\lambda_{.j} - x_{.j})}{\lambda_{1j}\,\lambda_{2j}(\lambda_{.j} - 1)}}}$$

which is based on Formula (13) and also finds frequent employment, differs only in weighting.

Criterion (16) was described by Cochran (3, p. 444). The formula is not strictly substantiated by him, although his principle of optimization (3, page 448-450) is essentially identical with ours.

Relevant examples of the application of this formula are found in Cochran (3, p. 442-446), as well as in Pearson (10) and Yates (15,16); the two authors mentioned last treat these examples by another method, however.